

Section 7 – Evaluation of Bias Report

INTRODUCTION

This report describes the results of an evaluation of the geocoding and mapping techniques for the Mapping Evaluation: Legal Services In Fast-Growing Counties In Southern California. The goals of the project are to assess the suitability and value of mapping techniques for legal services in Southern California.

A series of maps and charts depicting variables relevant to legal aid were provided for this project. For the evaluation of bias, a sampling of maps and charts were produced using techniques and methods developed from the previous mapping project in Georgia. In addition, a geocoding results dataset for statistical evaluation was provided. This evaluation of bias is to ensure that the cartographic techniques and geocoding methodology used is statistically robust, and accurately represents the data provided. In other words, checking the geocoding assignments to see if there was any systematic bias in their assignment that could effect the interpretation of the maps. Overall, the case geocoding rates displayed in Appendix B were very high for the Southern California project, decreasing the likelihood for bias.

CONCLUSIONS

There is no significant bias in the three client database fields tested. The departures from the mean are relatively small. The highest departures from the mean were found in the “none” category (no ethnicity indicated) of the “ethnicity” field. This reflects the case records where in addition to no ethnicity provided, no address or zip code information is provided; therefore neither geocoding method is possible. These are cases where small amounts of data were collected to document these cases in general.

DISCUSSION OF FINDINGS

Geocoding Location Bias

As covered in Section 6, two methods were applied to geocode the grantee database: MapMarker (geocoding software) and an “enhanced” geocoding method, developed to assign case records not successfully geocoded by the MapMarker method to a Block Group based on the ZIP Code area. Before evaluation of any potential geocoding bias within the grantee database, each resulting geocoded record was classified into five groups based on the geocoding results type:

- S – Single address match (MapMarker)
- M – Multiple address candidate match (MapMarker)
- Z – successful match at the ZIP code centroid level (MapMarker)
- EN - ENhanced geocoding method
- NG – Not Geocoded by either method

To evaluate geocoding bias, percentages and departures from the mean of geocoding match levels were compared within three client database fields: problem code, closure code and

ethnicity (Tables 7.1 – 7.12). Because there are a large number of problem and closure code categories, the top six categories (Table 7.0) were analyzed and a seventh category (“other”) incorporates all other problem and closure code records.

Table 7.0 – Problem and Closure Code Descriptions

Problem Code	Description
63	Private Landlord/Tenant
37	Spouse Abuse
32	Divorce/Separation
69	Other Housing
31	Custody/Visitations
51	Medicaid

Closure Code	Description
A	Counsel and Advise
B	Brief Services
I	Court Decision
F	Negotiated Settlement w/o Litigation
G	Negotiated Settlement w/ Litigation
H	Administrative Agency Decision

PROBLEM CODE

Table 7.1 – Problem Codes and Total Cases of Each Geocoding Assignment Category

Problem Code	M	S	Z	EN	NG
63	1751	25098	1788	227	104
37	280	9604	965	140	263
32	559	9258	709	273	122
69	394	7358	509	62	58
31	382	6477	643	121	165
51	316	5537	563	110	72
Other	1705	33495	2643	1027	736

Table 7.2 – Problem Codes and the Percentage of Each Geocoding Assignment Category

Problem Code	M	S	Z	EN	NG
63	6.04	86.64	6.17	0.78	0.36
37	2.49	85.35	8.58	1.24	2.34
32	5.12	84.77	6.49	2.50	1.12
69	4.70	87.79	6.07	0.74	0.69
31	4.90	83.17	8.26	1.55	2.12
51	4.79	83.92	8.53	1.67	1.09
Other	4.30	84.57	6.67	2.59	1.86
Mean	4.95	84.49	6.53	2.49	1.54

For each problem code, the geocoding assignment category [S, M, Z, EN, NG] percentages sum across each row to 100%, and reflect all cases being analyzed for each problem code. The mean

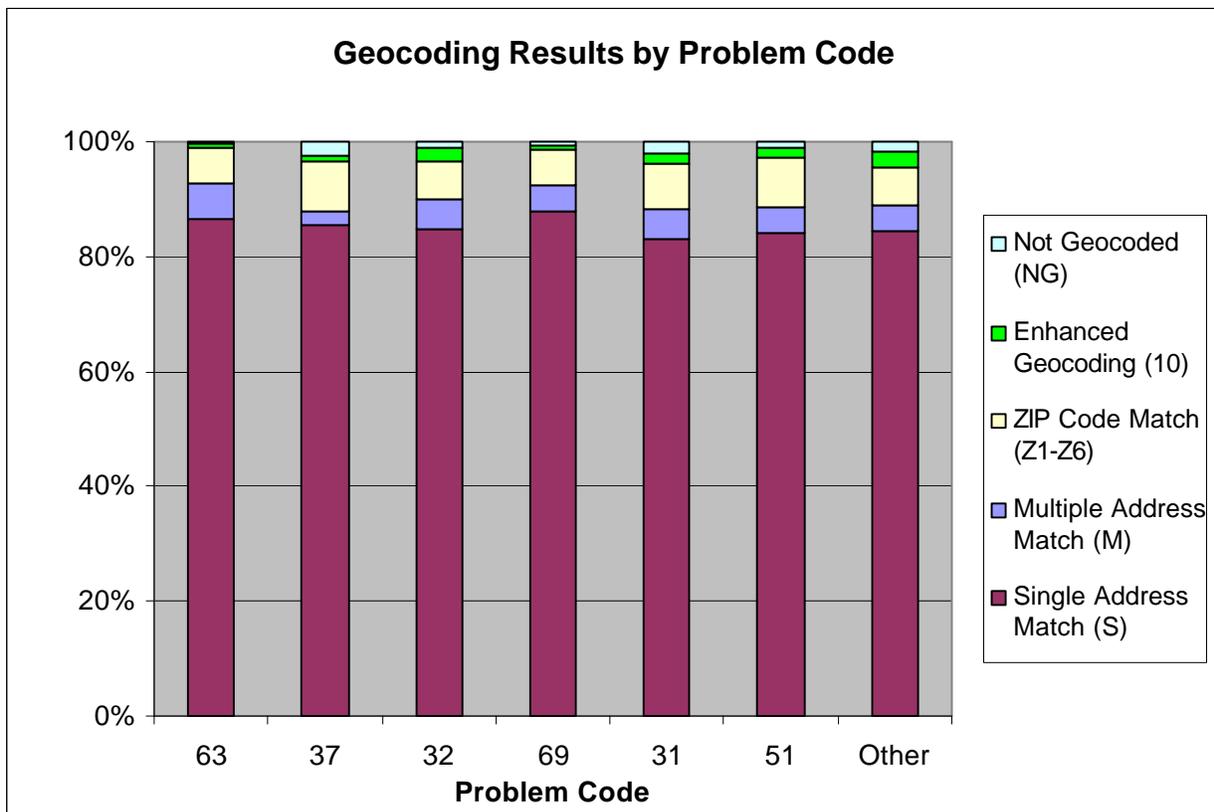
in the table above reflects the mean for all 113,514 addresses geocoded in this project, and not the mean for just the project codes displayed.

Table 7.3 – Departure from Mean of Table 7.2 Data

Problem Code	M	S	Z	EN	NG
63	1.10	2.15	-0.36	-1.70	-1.18
37	-2.46	0.87	2.04	-1.24	0.80
32	0.17	0.28	-0.04	0.01	-0.42
69	-0.25	3.31	-0.46	-1.75	-0.85
31	-0.04	-1.32	1.72	-0.93	0.58
51	-0.16	-0.57	2.00	-0.82	-0.45
Other	-0.64	0.08	0.14	0.11	0.32

To calculate the departure from the mean, the means in table 7.1 were subtracted from each of the assignment values in Table 7.1.

Table 7.4 – Geocoding Results by Problem Code



Note the similarity or consistency in the results.

CLOSURE CODE

Table 7.5 – Closure Codes and Total Cases of Each Geocoding Category Match

Closure Code	M	S	Z	EN	NG
A	2861	53061	4354	1170	758
B	1941	33238	2760	526	533
I	135	2324	134	40	82
F	152	2255	223	36	27
G	103	1911	83	33	11
H	61	1552	76	75	31
Other	134	2486	190	80	78

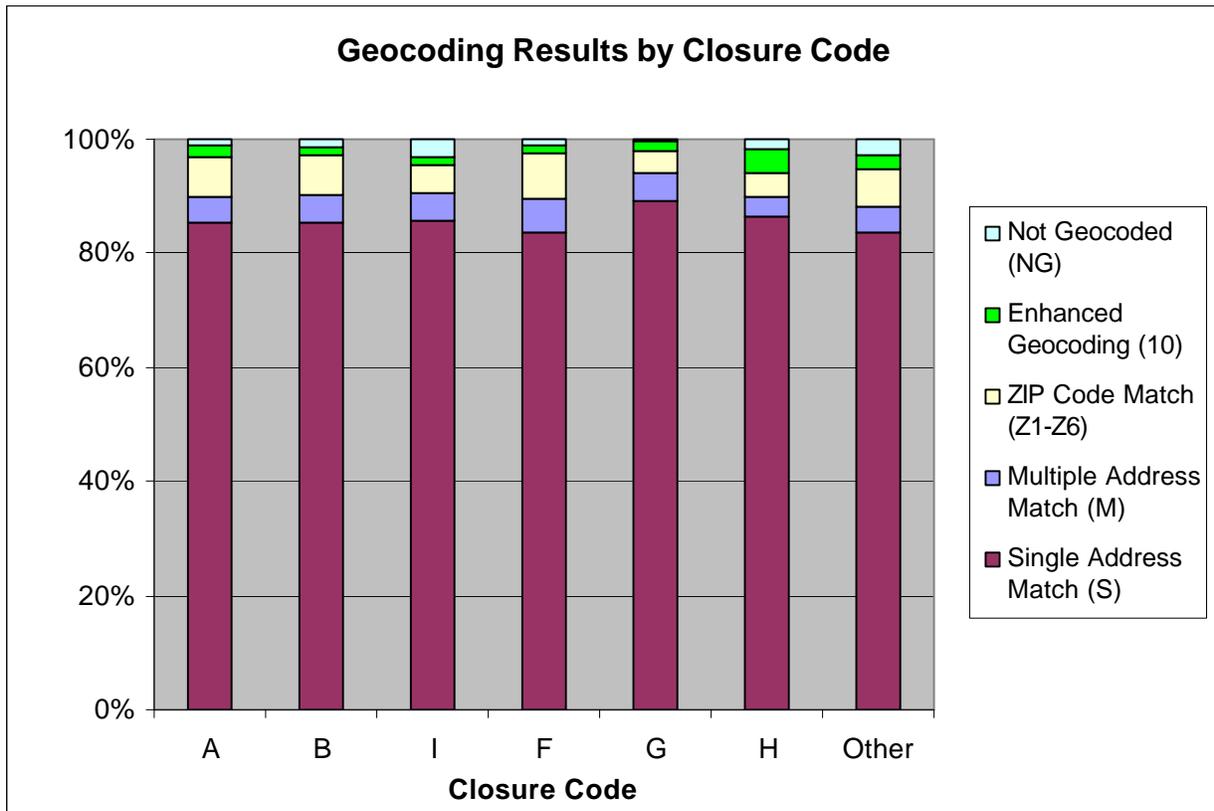
Table 7.6 – Closure Codes and the Percentage of Each Geocoding Category Match

Closure Code	M	S	Z	EN	NG
A	4.60	85.30	7.00	1.88	1.22
B	4.98	85.23	7.08	1.35	1.37
I	4.97	85.60	4.94	1.47	3.02
F	5.64	83.74	8.28	1.34	1.00
G	4.81	89.26	3.88	1.54	0.51
H	3.40	86.46	4.23	4.18	1.73
Other	4.51	83.76	6.40	2.70	2.63
Mean	4.61	84.93	6.30	2.28	1.89

Table 7.7 – Departure from Mean of Table 7.6 Data

Closure Code	M	S	Z	EN	NG
A	-0.01	0.37	0.70	-0.40	-0.67
B	0.37	0.30	0.78	-0.93	-0.52
I	0.37	0.67	-1.36	-0.81	1.13
F	1.04	-1.19	1.99	-0.95	-0.88
G	0.20	4.33	-2.42	-0.74	-1.37
H	-1.21	1.53	-2.06	1.90	-0.16
Other	-0.09	-1.17	0.11	0.41	0.74

Table 7.8 – Geocoding Results by Closure Code



ETHNICITY

Table 7.9 – Ethnicity and Total Cases of Each Geocoding Category Match

Ethnicity	M	S	Z	EN	NG
Hispanic	2446	41292	3579	502	548
White	1323	26754	1991	909	478
Black	1222	20961	1532	332	246
Asian	197	3757	358	82	113
Other	155	3203	285	98	91
Native American	38	714	56	33	17
None	6	146	19	4	27

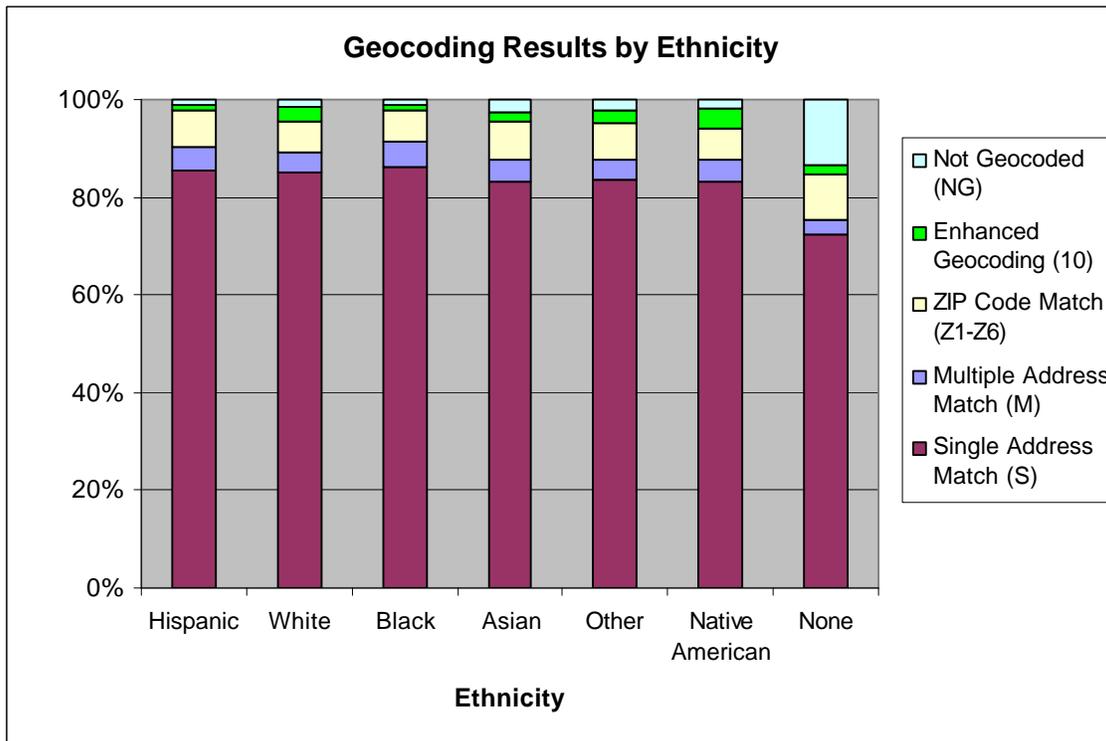
Table 7.10 – Ethnicity and the Percentage of Each Geocoding Category Match

Ethnicity	M	S	Z	EN	NG
Hispanic	5.06	85.37	7.40	1.04	1.13
White	4.21	85.05	6.33	2.89	1.52
Black	5.03	86.28	6.31	1.37	1.01
Asian	4.37	83.36	7.94	1.82	2.51
Other	4.04	83.59	7.44	2.56	2.37
Native American	4.43	83.22	6.53	3.85	1.98
None	2.97	72.28	9.41	1.98	13.37
Mean	4.30	82.74	7.34	2.21	3.41

Table 7.11 - Departure from Mean of Table 7.10 Data

Ethnicity	M	S	Z	EN	NG
Hispanic	0.76	2.64	0.06	-1.18	-2.28
White	-0.10	2.32	-1.01	0.68	-1.89
Black	0.73	3.55	-1.03	-0.85	-2.40
Asian	0.07	0.62	0.61	-0.39	-0.91
Other	-0.26	0.85	0.10	0.34	-1.04
Native American	0.13	0.48	-0.81	1.63	-1.43
None	-1.33	-10.46	2.07	-0.23	9.95

Table 7.12 – Chart of Geocoding Results by Ethnicity



Reporting Bias

Ten thematic maps were evaluated for reporting bias issues such as mapping non-normalized population data. All maps evaluated mapped population densities; therefore do not present this type of reporting bias.

The distribution of cases that were not successfully geocoded by either method was examined for geographic and address-type bias in the greater Los Angeles and Orange county area (Figure 7.1). Of particular interest were zip code areas with greater than ten percent of cases not successfully geocoded. As the map indicates, there are very few areas with not-geocoded percentages greater than five percent. An exception is in the Pasadena area with two zip codes that have a geocoding failure rate of 20-25%. Because the total number of cases in these zip codes is 17, a geocoding failure rate of this percentage is probably not due to any geographic bias.

Table 7.13 illustrates the totals and percentages of cases not geocoded by address type. The majority of these address types have either no or incomplete address information provided, or are homeless/shelter cases. An overall geocoding success rate of 98.66% out of 113,514 total cases indicates low potential for any significant bias.

Figure 7.1 – Map of the distribution of cases not geocoded by Zip code in the greater Los Angeles area.

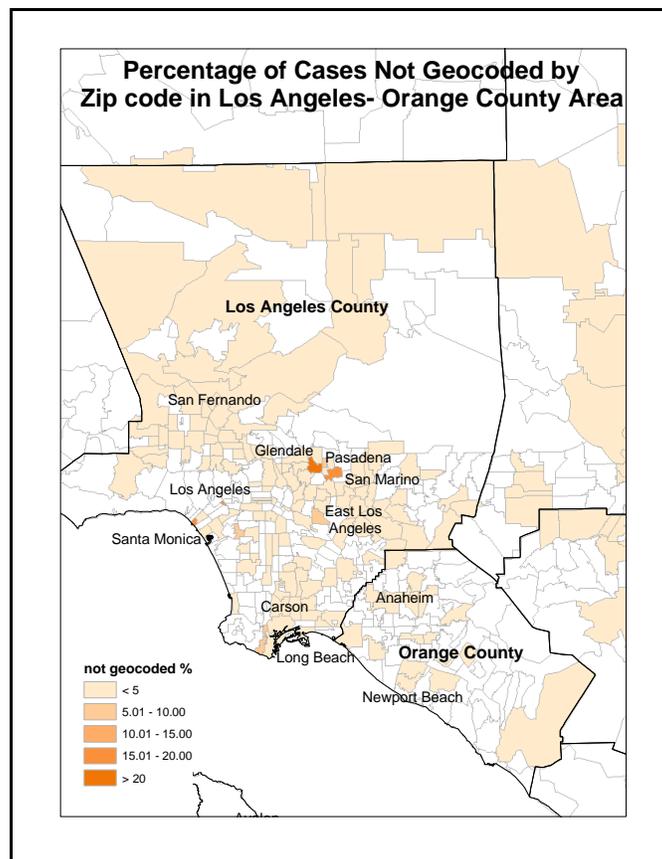


Table 7.13 – Cases Not Geocoded by Address Type

Address type	Not Geocoded	Total cases
Correctional	1	1
Institutional	18	18
PO Box	36	1,996
C/O	49	49
Out of State	122	122
Shelter	203	203
Not Provided	205	205
Insufficient	273	273
Homeless	291	291
Residential	322	110,356
Total	1,520	113,514

It should be noted the reason none of the Shelter cases were geocoded was to preserve the confidentiality of shelter locations, which is a different rationale than not geocoding the homeless because they don't have an address.